

ALGORITHM AND BENCHMARK DATASET FOR STAIN SEPARATION IN HISTOLOGY IMAGES

Michael T. McCann¹, Joshita Majumdar¹, Cheng Peng³, Carlos A. Castro⁴, Jelena Kovačević^{2,1}

¹Dept. of BME and Center for Bioimage Informatics, ²Dept. of ECE, Carnegie Mellon University, Pittsburgh, PA, USA

³PSsolution, Pittsburgh, PA, USA

⁴Dept. of Obstetrics and Gynecology,

Magee-Womens Research Institute and Foundation of the University of Pittsburgh, Pittsburgh, PA, USA

ABSTRACT

In this work, we present a new algorithm and benchmark dataset for stain separation in histology images. Histology is a critical and ubiquitous task in medical practice and research, serving as a gold standard of diagnosis for many diseases. Automating routine histology analysis tasks could reduce health care costs and improve diagnostic accuracy. One challenge in automation is that histology slides vary in their stain intensity and color; we therefore seek a digital method to normalize the appearance of histology images. As histology slides often have multiple stains on them that must be normalized independently, stain separation must occur before normalization. We propose a new digital stain separation method for the universally-used hematoxylin and eosin stain; this method improves on the state-of-the-art by adjusting the contrast of its eosin-only estimate and including a notion of stain interaction. To validate this method, we have collected a new benchmark dataset via chemical destaining containing ground truth images for stain separation, which we release publicly. Our experiments show that our method achieves more accurate stain separation than two comparison methods and that this improvement in separation accuracy leads to improved normalization.

Index Terms— histology, unmixing, stain separation

1. INTRODUCTION

Histology, the study of the microscopic structure of tissues, is a critical tool in medical research and serves as the gold standard for the diagnosis of many diseases. Histological stains are used to highlight the detail in tissues and cells. In the case of the inexpensive and common hematoxylin and eosin (H&E) stain, hematoxylin stains cell nuclei blue and eosin stains other structures pink or red.

Current research (see [1] for a review) aims to automate some aspects of histology analysis, but variations in stain color and intensity between slides complicate this problem. These variations may arise from differences in staining protocol, differences between stain brands, or the shelf life of stains. Thus, automated histology image analysis would benefit from accurate stain normalization techniques. Because normalization is difficult as images contain different stains, stain separation is an important first step in stain normalization. To

address this need, this work presents a new algorithm and benchmark dataset for digital stain separation.

2. PREVIOUS WORK

There exist a variety of approaches to digital stain separation for histology images in the literature. Some of these methods are unsuitable for our purposes because they place special restrictions on the imaging setup used. For example, [2] requires knowledge of the stain colors used and is therefore only useful when such calibration data is present. The method in [3] uses non-negative matrix factorization (NMF) and does not require calibration data, but relies on collecting hyperspectral images using expensive and non-standard imaging equipment.

Some authors have addressed the separation of stains in RGB images of H&E-stained slides without calibration, as is our focus here. For example results from these methods as well as our proposed method, see Figure 1. In [4], stain separation is based on the idea that the contrast in the red channel comes from hematoxylin. Therefore, the method finds the E-only image by adding an inverted and scaled version of the image’s red channel to its red, green, and blue channels, searching for the scaling in each channel that reduces its variance the most. The result is an image with constant red channel and low variance in the green and blue channels. These changes make the image mostly red and lacking in nuclei, so this is the E-only estimate. The H-only estimate is found by subtracting the E-only estimate from the input and re-white balancing. Throughout the paper, we will refer to this method as the *color inversion* method.

We, instead, build on the work in [5]; it is explained in detail in Section 4.1. The idea is that the color of each pixel in the H&E image is a positive linear combination of the two unknown stain vectors. This means that all the pixels of the image lie inside a wedge with its apex at the color white and its edges determined by the color of hematoxylin and eosin; for this reason we call this approach *wedge finding*. Wedge finding is a special case of NMF, but rather than rely on generic algorithms, the authors propose a geometric solution which they show to be more efficient than the NMF technique in [3].

3. BENCHMARK DATASET

A key obstacle in advancing the state of the art in digital stain separation is the lack of a publicly available benchmark dataset. In [4], a benchmark is created via a process of staining and destaining a single slide; but it is only used for qualitative evaluation of the result and is not, to our knowledge, available online.

Cheng Peng performed this research while at ¹. The authors gratefully acknowledge support from the National Science Foundation Graduate Research Fellowship under Grant No. 0946825, the Philip and Marsha Dowd Teaching Fellowship, the John and Claire Bertucci Fellowship, the Pittsburgh Chapter of the Achievement Rewards for College Scientists Foundation Scholarship, and the CMU Carnegie Institute of Technology Infrastructure Award. Email: mtmccann@cmu.edu

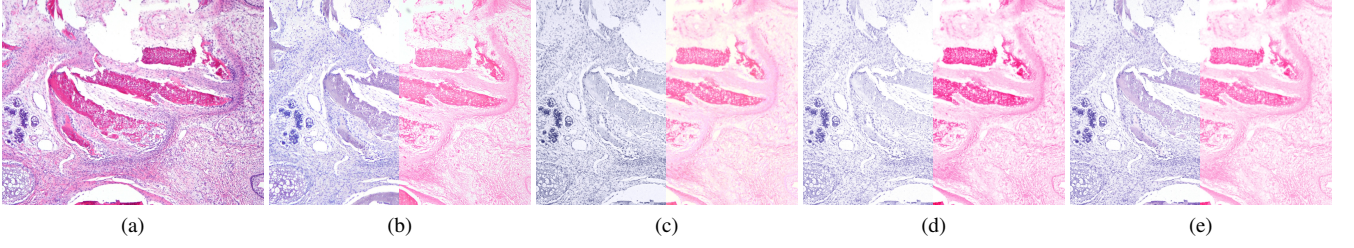


Fig. 1. Example stain separation showing (a) an input image of an H&E-stained slide, (b) the ground truth stain separation, and the digital stain separation results from (c) color inversion [4], (d) wedge-finding [5], and (e) our proposed method. The left half of each image shows the H-only estimate and the right half shows the E-only estimate.

Following a procedure similar to that of [4], we have created our own stain separation benchmark using chemical destaining according to standard histological practice [6]. Our goal was to generate images of the same slide under three staining conditions: H-only, E-only, and H&E. Thus, the H-only and E-only images can serve as the ground truth for separation of the H&E image.

Our procedure was as follows. (1) Image an unstained slide. (2) Apply eosin stain and image. (3) Destain the slide totally and image. (4) Apply hematoxylin stain and image. (5) Apply eosin stain and image. At each imaging step, we photographed the slide at the same three locations, giving us three images of the same slide under blank, E-only, destained, H-only, and H&E conditions (We did not capture more images because they are essentially redundant as we don't expect the stain to vary spatially on the slide). We show examples of the H-only, E-only, and H&E slides in Figure 1. Although we do not show them for space, the blank and destained images serve as controls. Since both show no staining, we can be certain that the destaining process was successful. We provide all of these images in the reproducible research compendium to this paper [7].

4. METHODS

In this section, we review the details of the wedge-finding method of [5], describe our proposed method, which extends it, and explain how we turn a stain separation technique into a stain normalization technique.

4.1. Wedge Finding [5]

Let f_{HE} be an $M \times N$ RGB image of an H&E-stained slide with the value $[1, 1, 1]^T$ corresponding to white and $[0, 0, 0]^T$ to black. Making the assumption that the stains follow the Beer-Lambert law, we can write the absorbance of the H&E image at each pixel x as a weighted sum of the unknown absorbances of the hematoxylin and eosin stains,

$$f_{\text{HE}}^{OD}(x) = -\log(f_{\text{HE}}(x)) = \varphi_H a_H(x) + \varphi_E a_E(x), \quad (1)$$

where φ_H and φ_E are unit-length 3×1 vectors giving the color of hematoxylin and eosin and $a_H(x)$ and $a_E(x)$ are positive scalars for all x . We refer to f_{HE}^{OD} as the optical density (OD) space version of f_{HE} , because high values correspond to high density of stain on the slide (dark colors).

Stacking all of the pixels of f_{HE}^{OD} together into a matrix, we can rewrite (1) as

$$f_{\text{HE}}^{OD} = \Phi_{\text{HE}} a_{\text{HE}},$$

$\begin{matrix} [3 \times MN] & [3 \times 2] [2 \times MN] \end{matrix}$

where $\Phi_{\text{HE}} = [\varphi_H \quad \varphi_E]$ and $a_{\text{HE}} = [a_H^T \quad a_E^T]^T$.

Performing stain separation on f_{HE}^{OD} then amounts to a NMF problem where the goal is to find Φ_{HE} and a_{HE} . To do this, we use singular value decomposition (SVD) to project the OD space image onto a plane,

$$f_{\text{HE}}^{OD} \approx \Phi \begin{matrix} [3 \times 2] \\ [2 \times MN] \end{matrix} a = [\varphi_0 \quad \varphi_1] \begin{bmatrix} a_0 \\ a_1 \end{bmatrix}.$$

To move from the arbitrary orthogonal basis Φ to an estimate of the H&E basis $\hat{\Phi}_{\text{HE}}$, we assume that some pixels exist that are stained with hematoxylin only and that some pixels exist that are stained with eosin only; therefore all the pixels should exist inside a wedge, the extreme members of which correspond to pure hematoxylin and pure eosin. The extreme pixels are those that form the maximum and minimum angle with the φ_0 axis, and $\hat{\varphi}_H$ and $\hat{\varphi}_E$ are their projections back into OD space,

$$\begin{aligned} \hat{\varphi}_H &= \Phi \begin{bmatrix} a_0(x^\Delta) \\ a_1(x^\Delta) \end{bmatrix} & x^\Delta &= \arg \max_x \arctan \left(\frac{a_1(x)}{a_0(x)} \right), \\ \hat{\varphi}_E &= \Phi \begin{bmatrix} a_0(x^\nabla) \\ a_1(x^\nabla) \end{bmatrix} & x^\nabla &= \arg \min_x \arctan \left(\frac{a_1(x)}{a_0(x)} \right). \end{aligned}$$

To complete the stain separation, we find weights, $\hat{a}_{\text{HE}} = [\hat{a}_H^T \quad \hat{a}_E^T]^T = (\hat{\Phi}_{\text{HE}})_{\text{left}}^{-1} f_{\text{HE}}^{OD}$, and project the input H&E image separately onto the hematoxylin and eosin vectors, so $\hat{f}_H^{OD} = \hat{\varphi}_H \hat{a}_H$ and $\hat{f}_E^{OD} = \hat{\varphi}_E \hat{a}_E$. Finally we convert back from OD space to RGB space, and reshape the results into the single-stain images \hat{f}_H and \hat{f}_E .

In practice, we replace the minimum and maximum with the α th and $(100 - \alpha)$ th percentile, which provides robustness against outlier pixels (e.g. dirt on the slide). We also exclude nearly-white pixels from these calculations because they become unstable when a_0 or a_1 is small. Finally, we may need to swap the identities of $\hat{\varphi}_H$ and $\hat{\varphi}_E$ if we find them to be the wrong colors, since the method is agnostic to the fact that the hematoxylin stain should be blue and the eosin stain pink.

4.2. Proposed Method

Our proposed method uses two observations from our benchmark dataset to build on the wedge-finding method presented above. First, real E-only images exhibit lower contrast when compared to digitally separated E-only images using the wedge-finding method (compare the right half of Figure 1b and Figure 1d to see this). Second, the optical density of some regions of the H&E image is significantly higher than the sum of the optical densities of the corresponding H-only and E-only image. This points to an interaction between the stains beyond summing in the OD space. Our proposed method is motivated by these observations.

Contrast of Eosin. Our first improvement to the wedge-finding method is to reduce the contrast of the E-only image. To better quantify the observation that real E-only images have low contrast, we examined the order statistics of the norms of the pixel values of our ground truth single-stain images in OD space. We found that the order statistics change smoothly and that for each pair of H- and E-only images, the order statistics of the two single-stain images were similar except for a multiplicative constant, $\|f_E^{OD}(x_i)\| \approx C\|f_H^{OD}(y_i)\|$, where x_i denotes the pixel location of the i th smallest element in f_E^{OD} and y_i denotes the pixel location of the i th smallest element in f_H^{OD} . Performing the same investigation on H- and E-only images created digitally using the wedge-finding method, we noted significant deviation from the smooth behavior for the E-only estimate.

Therefore, starting with \hat{a}_{HE} found via the wedge-finding method, we estimate C as the median of $\hat{a}_E(x_i)/\hat{a}_H(y_i)$. To correct the contrast of the E-only image, we let $\hat{a}_E^*(x_i) = (1 - \gamma)\hat{a}_E(x_i) + \gamma C\hat{a}_H(y_i)$, where γ controls the degree of smoothing: when $\gamma = 0$, the contrast reduction has no effect; when $\gamma = 1$, the eosin coefficients are shifted so their order statistics are exactly the same as the hematoxylin coefficients times a constant. We can then compute a new stain separation

$$\hat{f}_E^{OD} = \hat{\varphi}_E \hat{a}_E^*, \quad \hat{f}_H^{OD} = \hat{f}_{HE}^{OD} - \hat{f}_E^{OD}.$$

Stain Interaction. The method described in the previous section results in improved estimates of the E-only image and, when γ is low, it also improves the H-only estimate by adding some pink which is indeed present in the ground truth (Figure 1b). When γ is high, the H-only estimate begins to suffer because it contains too much pink; the previous model in (1) does not account for this. According to (1), whenever the E-only estimate improves, the H-only estimate should too, because they exactly sum to the H&E image. We therefore extend that model by adding an interaction term,

$$f_{HE}^{OD}(x) = \varphi_H a_H(x) + \varphi_E a_E(x) + I(x).$$

In general, $I(x)$ could have a complicated dependence on $a_H(x)$, $a_E(x)$, and the underlying tissue, making it hard to calculate directly. Instead, we assume that interaction occurs only in those places where the contrast of the eosin coefficients is adjusted. We let some fraction λ of the change in eosin be due to that interaction and the rest be due to pinkness within the H-only image, giving our final stain separation expression,

$$\hat{f}_E^{OD} = \hat{\varphi}_E \hat{a}_E^*, \quad \hat{f}_H^{OD} = (1 - \lambda)(\hat{f}_{HE}^{OD} - \hat{f}_E^{OD}) + \lambda \hat{\varphi}_H \hat{a}_H.$$

When λ is high, most of the adjustment in the E-only image is assumed to be due to the interaction term, so \hat{f}_H^{OD} is unaffected; when λ is low, the eosin adjustment is assumed to be due to pinkness in the H-only image, so \hat{f}_H^{OD} changes a lot.

4.3. Stain Normalization

We use the same normalization scheme for each method, which is similar to the approach taken in [5]. Given two H&E images, our goal is to normalize the input image to look like the target image. To do this, we stain separate both input and target images and represent each (OD space) pixel in the single-stain estimates as a scalar weight times a unit-length stain vector. For the color inversion technique [4], we find the stain vectors and weights via 1D SVD on the single-stain images. For the wedge-finding method [5], the stain vectors are $\hat{\varphi}_H$ and $\hat{\varphi}_E$ and the weights are \hat{a}_H and \hat{a}_E . For

Table 1. Results of the stain separation comparison reported as SNR [dB] for hematoxylin and eosin, with the best result for each image and stain in bold.

	Method					
	Color inversion [4]		Wedge finding [5]		Proposed	
	H	E	H	E	H	E
Image 1	26.13	25.54	26.07	25.34	27.31	26.42
Image 2	23.39	22.89	23.29	21.54	23.56	23.04
Image 3	25.20	26.04	28.48	24.89	29.56	27.32
Mean	24.91	24.82	25.95	23.92	26.81	25.59

our method, we also use $\hat{\varphi}_H$ and $\hat{\varphi}_E$ as the stain vectors, but use \hat{a}_E^* for the eosin weights and $\|\hat{f}_H^{OD}(x)\|$ for the hematoxylin weight at each pixel x . To adjust for variations in stain intensity, we independently scale the input eosin and hematoxylin weights so that their 95th percentile values match the 95th percentile values of the corresponding weights from the target image. To adjust for variations in stain color, we replace the input stain vectors with those of the target. We complete the normalization process by recombining the single-stain images of the input.

5. EXPERIMENTS AND RESULTS

We validate our method with three experiments: (1) We compare our stain separation method to the color inversion [4] and wedge finding [5] methods quantitatively, using our the newly-created benchmark dataset (described in Section 3). (2) We again use the benchmark dataset to explore the effects of our method’s three parameters. (3) We compare the performance of the same three methods on a stain normalization task. We now describe these experiments and their results in detail.

5.1. Stain Separation Comparison

For each of the three methods, we performed a stain separation on each of the three images H&E images in our dataset. We set parameters at typical values: for the wedge finding, $\alpha = 0.5$; for our method, $\alpha = 0.5, \gamma = .75, \lambda = 0.5$. We explore parameter selection in depth in the next section; here our goal was to compare the methods with default parameters, since in the real use case for stain separation no ground truth exists with which to tune parameters. To quantitatively evaluate the results, we use affine registration to align each ground truth H- or E-only image with the corresponding H&E image. Then, for each digital separation result, we blur the ground truth image and the result with a Gaussian ($\sigma = 15\text{px}$) and calculate the signal to noise ratio (SNR); this blurring is intended to reduce the effect of small misalignments between the two images being compared, it does not affect the trend in the results.

Figure 1 shows the qualitative results on a single image and Table 1 shows the results in terms of SNR. Both our method and the color inversion method result in low contrast for the E-only estimate; but our method additionally captures some of the pink tone in the H-only estimate, especially in the bone tissue (long regions in the center of the image). Quantitatively, our proposed method outperforms the others for each image and stain.

5.2. Parameter Analysis

In order for our method to be really useful, its strong performance must hold over a wide range of parameters. To explore whether this

is true, we performed a parameter analysis over the wedge percentile α , the eosin contrast reduction γ , and the interaction amount λ for the first image in our dataset (first row of Table 1).

We began by investigating α and γ with λ set to 0 (Figure 2a). Focusing on the $\gamma = 0$ case (solid lines), we see that α does not affect performance much unless it is too small; a setting of 0.5 or slightly less is optimal for this image, but larger values also work. This is intuitive because setting α to be very small would allow outliers to greatly affect the estimation of the hematoxylin and eosin vectors. Increasing γ (from solid to dashed to dotted lines) always improves the estimate for the E-only image, but at the cost of the H-only estimate for high values. Intuitively, the H-only image gets too red when γ is high. The mean SNR shows that $\gamma = 1$ (dotted line) is still a good choice for all values of α .

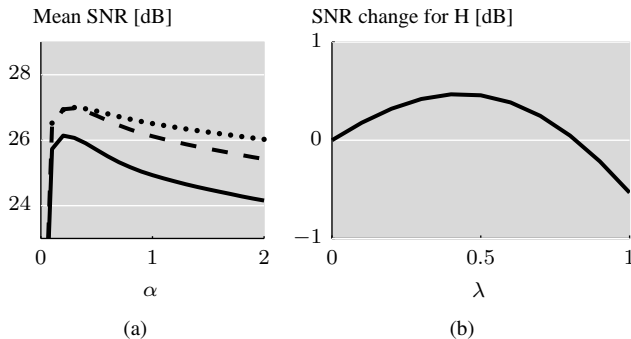


Fig. 2. Results of parameter experiments. In (a), the effects of wedge percentile α and eosin contrast reduction γ (solid, dashed, dotted lines corresponding to $\gamma = 0, 0.5, 1$) on stain separation accuracy. When $\gamma = 0$, our method reduces to wedge finding [5]; increasing γ improves the mean SNR. In (b), the effects of interaction amount λ for wedge percentile $\alpha = 0.3$ and eosin contrast reduction $\gamma = 1$. A wide range of λ values all improve the SNR of the H-only estimate.

Next, we explore how increasing the interaction amount λ can improve the estimate of the H-only image. Starting at the point that gives highest mean SNR ($\alpha = 0.3$, $\gamma = 1$) in the previous experiment, we vary λ (Figure 2b). For any λ between 0 and 0.8, the SNR of the H-only estimate improves, with a maximum improvement of .47dB at $\lambda = 0.4$. Taken together, these parameter studies show that the proposed method is robust with respect to parameter choices.

5.3. Stain Normalization Comparison

We now compare the performance of the stain separation methods on a normalization task. For each method, we used the process described in Section 4.3 to normalize an input image (Figure 1a) to a target image (Figure 3a). Qualitatively, the normalization results (Figure 3) are similar among the wedge-finding, color inversion, and our proposed method.

Quantifying the results of this experiment is difficult because simply subtracting the normalized and target images yields no useful information, since the images are of different tissue areas. We therefore seek an image similarity metric that compares the colors occurring in each image without accounting for their location or frequency. To accomplish this, we generate a signature (along the lines of [8]) for the normalized and target image by clustering their pixels using k-means in the Lab color space, which we selected for its perceptual uniformity. We then calculate an optimal linear assignment

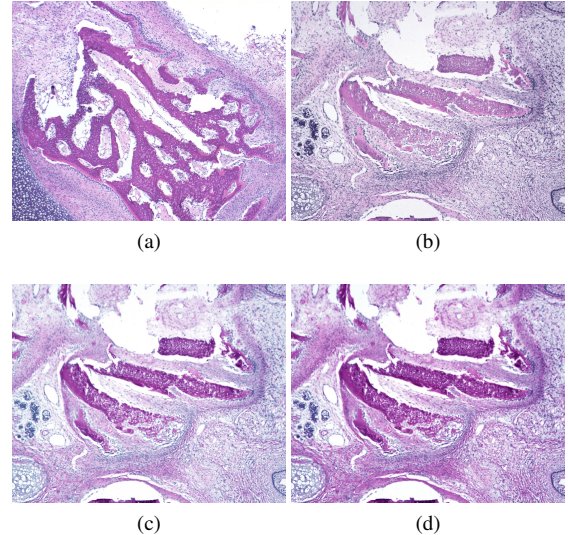


Fig. 3. Results of the stain normalization experiment, showing (a) the target and (b) the results using color inversion [4], (c) wedge-finding [5], and (d) our proposed method. The results are similar across methods; the main difference is the depth of color in the bone tissue (long purple structure in (b)-(d)).

(with the MATLAB code [9]) between the sets of cluster centers. We use the total cost of the optimal assignment as our measure of distance. Using ten cluster centers, we find the total costs of assignment to be 122.4, 96.3, and 51.1 for color inversion, wedge-finding, and our method, respectively. Although this experiment is small and the proper distance metric is an open question, these results at least suggest that our method’s improved stain separation performance translates into more accurate stain normalization as well.

6. CONCLUSIONS AND FUTURE WORK

We have presented a new algorithm and benchmark dataset for digital H&E stain separation. The benchmark provides a way to quantitatively evaluate digital stain separations via images of chemically destained slides. We showed that our algorithm outperforms the currently-available methods on all test images and is robust to parameter selection. We also showed that this improved separation leads to improved stain normalization.

In our future work, we will explore more principled priors on the E-only contrast, perhaps by finding a parametric description of the order statistics. We would also like to continue to improve the forward model. One approach would be to explore more complex lighting models including reflection in addition to absorbance; another would be to study the interaction term more closely. All of this work would benefit from chemical experiments including a wider variety of stain brands, staining procedures, lighting conditions, and tissues. As we collect this data, we will continue to make it publicly available.

7. REFERENCES

- [1] M. Gurcan, L. Boucheron, A. Can, A. Madabhushi, N. Rajpoot, and B. Yener, "Histopathological image analysis: A review," *IEEE Rev. Biomed. Eng.*, vol. 2, pp. 147–171, Apr. 2009.
- [2] A. C. Ruifrok and D. A. Johnston, "Quantification of histochemical staining by color deconvolution," *Analytical and quantitative cytology and histology / the International Academy of Cytology [and] American Society of Cytology*, vol. 23, no. 4, pp. 291–299, Aug. 2001.
- [3] A. Rabinovich, S. Agarwal, C. A. Laris, J. H. Price, and S. Be-longie, "Unsupervised color decomposition of histologically stained tissue samples," in *In Advances in Neural Information Processing Systems*. Suntech City: MIT Press, 2003.
- [4] P. Tadrous, "Digital stain separation for histological images," *Journal of Microscopy*, vol. 240, no. 2, pp. 164–172, 2010.
- [5] M. Macenko, M. Niethammer, J. Marron, D. Borland, J. Woosley, X. Guan, C. Schmitt, and N. Thomas, "A method for normalizing histology slides for quantitative analysis," in *Proc. IEEE Int. Symp. Biomed. Imag.*, Chicago, IL, Jun. 2009, pp. 1107–1110.
- [6] G. W. Gill and D. M. Wood, "Biological microtechniques," *DAKO Connection*, vol. 14, pp. 144–150, 2010.
- [7] M. T. McCann, J. Majumdar, C. Peng, C. A. Castro, and J. Kovačević. (2014, Jan.) Algorithm and benchmark dataset for stain separation in histology images. [Online]. Available: http://www.jelena.ece.cmu.edu/repository/rr/14_McCannMPCK/14_McCannMPCK.html
- [8] C. Beecks, M. Uysal, and T. Seidl, "A comparative study of similarity measures for content-based multimedia retrieval," in *Multimedia and Expo (ICME), 2010 IEEE International Conference on*, 2010, pp. 1552–1557.
- [9] Y. Cao. (2014, Jan.) Hungarian algorithm for linear assignment problems (v2.3). [Online]. Available: <http://www.mathworks.com/matlabcentral/fileexchange/20652-hungarian-algorithm-for-linear-assignment-problems-v2-3>